**CS231n学习报告**

1. **Image Classification with Linear Classifiers：**
2. Nearest Neighbor

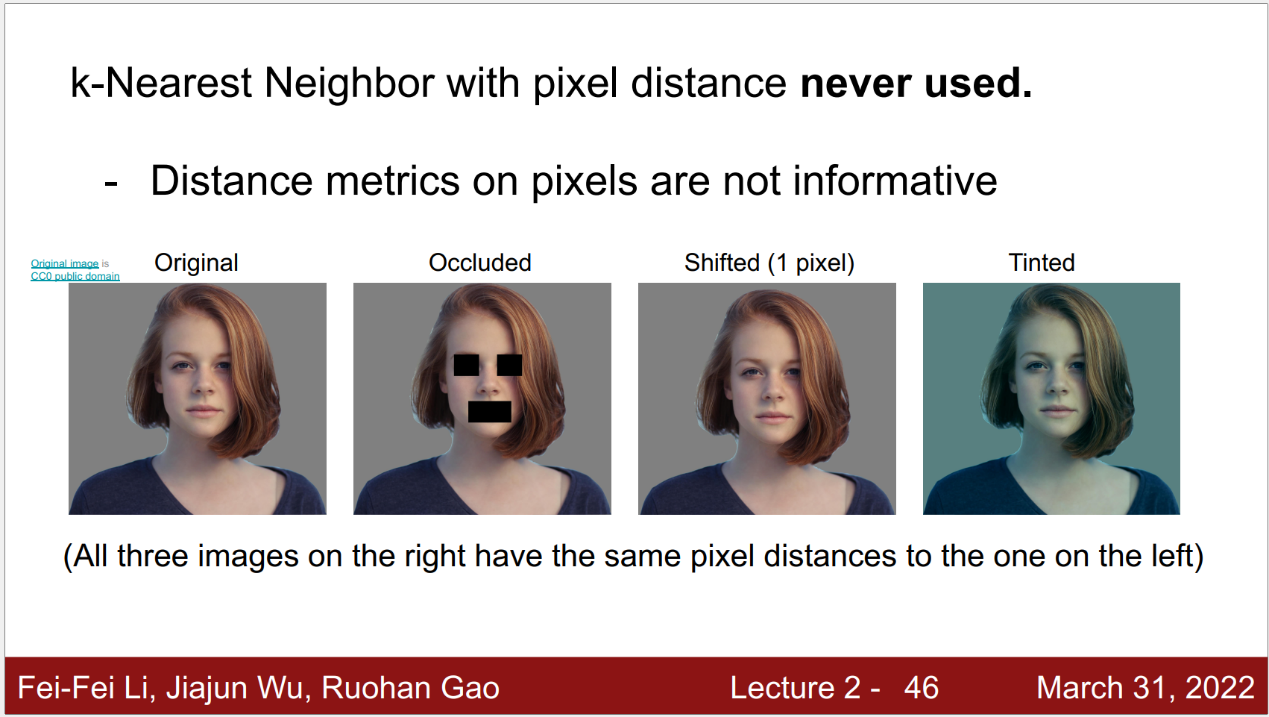
2、K-Nearest Neighbors

【Cross-Validation：Split data into folds, try each fold as validation and average the results（for small datasets）】

【slow at test time】

【像素间距离在不适用于表示图像距离，不能表示图像在视觉上上的相似度。】

**（视觉上相似的图片在像素距离上可能相差很大，后面数据增强也是解决这个问题，但是为什么ppt上说右三张与左边原图有着相同的像素距离？）**



**【**Curse of dimensionality**】**

* 1. Distance Metrics：

a) Minkowski Distance：



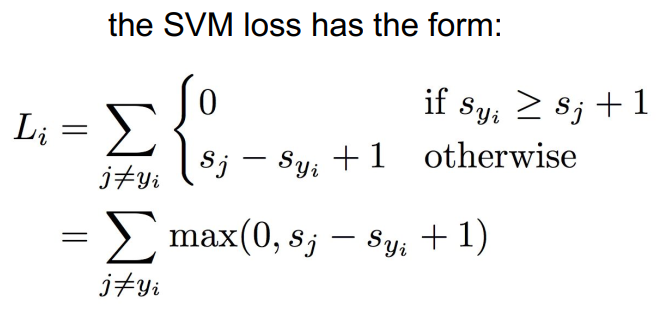
b) Mahalanobis Distance：

(：该分布的协方差矩阵)

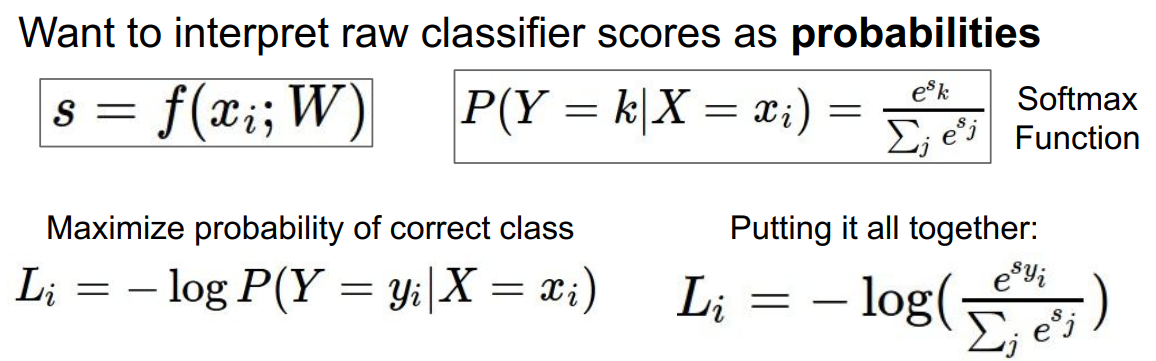
c) Metric Learning：**度量学习在CV的模型中的应用？**

4、linear classification：

* + 1. Multiclass SVM loss



* + 1. Softmax classifier / Multinomial Logistic Regression



5、classification <-> regression：回归（连续值）经过激活函数变为分类（离散值）

**二、Regularization：**

正则化，解决过拟合

1、L1很可能会得到只有一个系数（坐标轴上），可以筛选对结果贡献最大的系数，但不

很稳定

2、L2比较稳定，多个系数综合

3、Dropout(随机失活)：一般情况下，对于同一组训练数据，利用不同的神经网络训练之后，求其输出的平均值可以减少overfitting

4、Data augmentation数据增强：

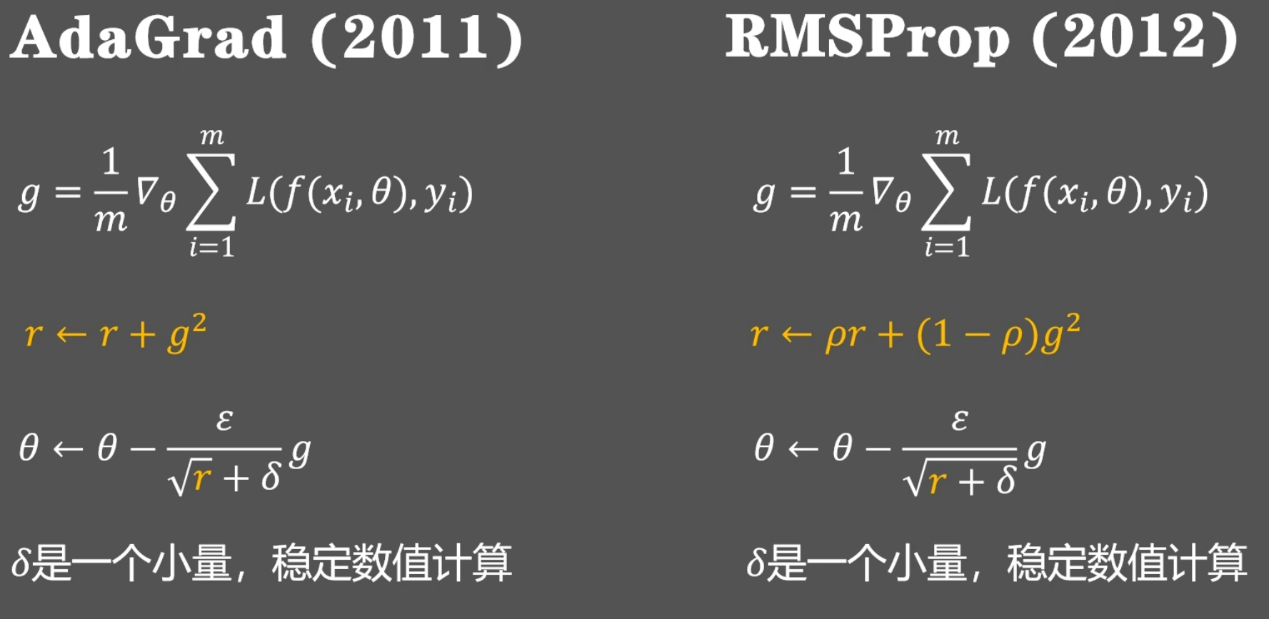
Flip镜像、rotation旋转、scale缩放、crop裁剪、translation平移、gaussian noise高斯噪声

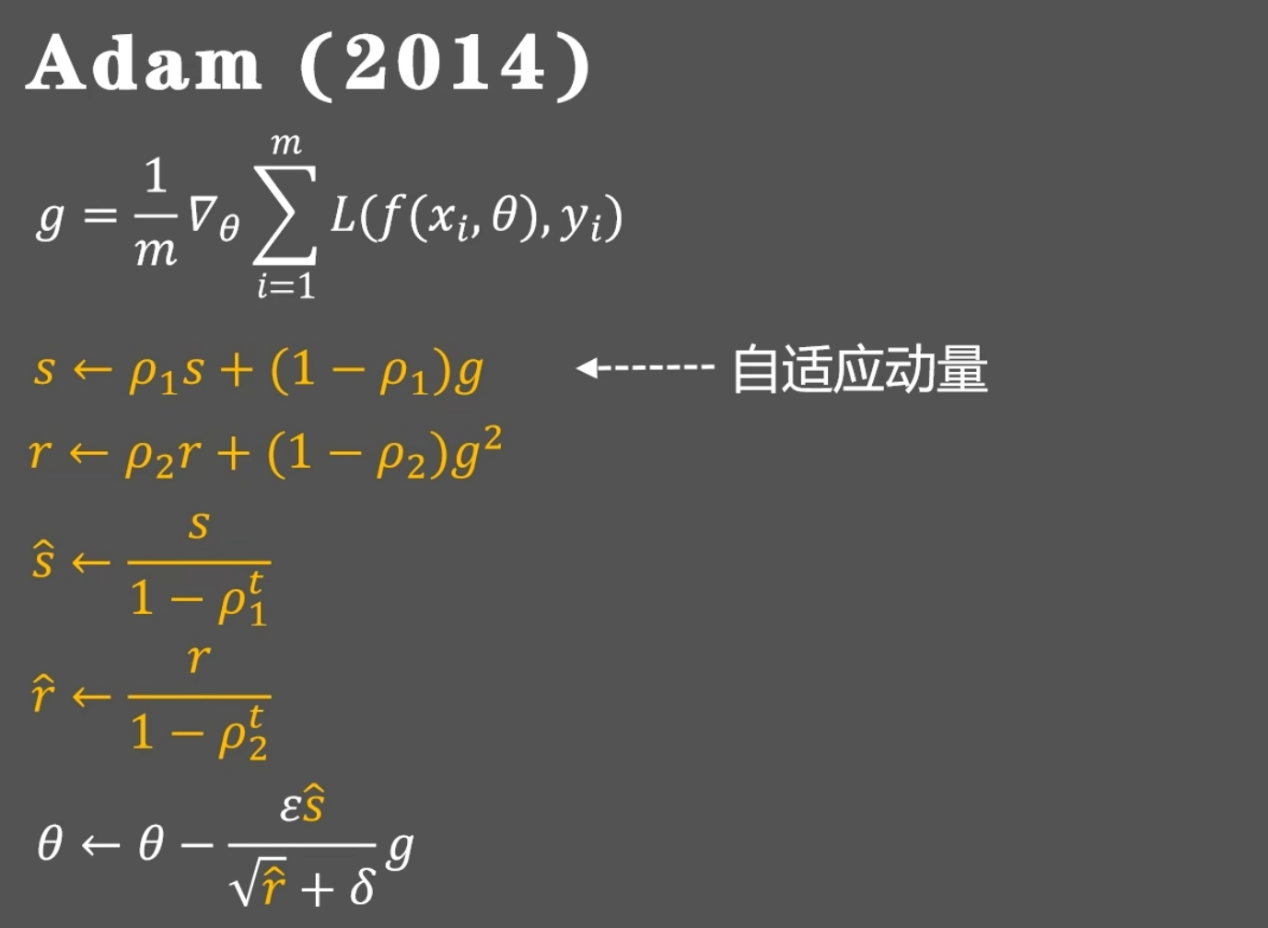
5、Cutout / Random crop

**三、最小化损失函数：**

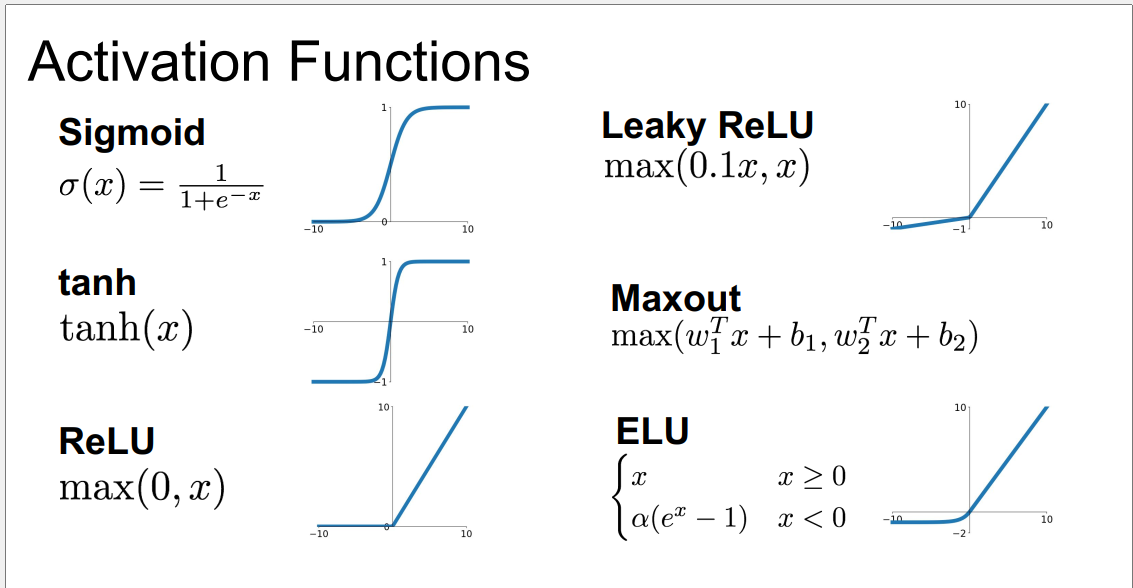
1、Gradient Descent（GD）：每次需要用到所有数据，更新速度慢 **->** Mini-batch gradient descent和stochastic gradient descent（SGD）： 每次只取小部分样本更新梯度，但是可能导致始终在最优点附近来回震荡，无法收敛 **->** gradient descent with momentum：使用动量（保留上一次的梯度方向）

2、Adagrad（2011）：通过梯度也把学习率调整，但可能学习率过早变小 **->** RMSProp **->** Adam（2014）（动量+学习率自适应）





3、梯度下降中使用饱和的非线性函数（sigmoid、tanh）会比非饱和的非线性函数（ReLU）慢很多，但是ReLU输出需要归一化



**四、 Logistic Regression：**

1、Linear Regression + Sigmoid （Linear Regression的输出作为Sigmoid输入） （pyspark库）

**五、 Optimization：**

1、牛顿法：二阶可导，Hesse矩阵求逆的计算复杂 -> 拟牛顿法：构造一个矩阵 U逼近H-1（Hessian Matrix）：DFP、BFGS、L-BFGS

（主要为**Convex-Optimization**的内容）

**六、Multilayer Perceptron（MLP）/ Artificial Neural Network（ANN）：**

**线性可分问题，学习过程一定会收敛的证明？（Minsky and Papert, 1969）**

**七、Computational graphs：**

Upstream gradient、local gradient、downstream gradient

downstream gradient=local gradient \* Upstream gradient

**八、vector derivatives (涉及向量函数微分学、矩阵函数)**

**九、batch normalization：**不预处理会使绝对值较大的数据不敏感（到了激励函数的饱和阶段）

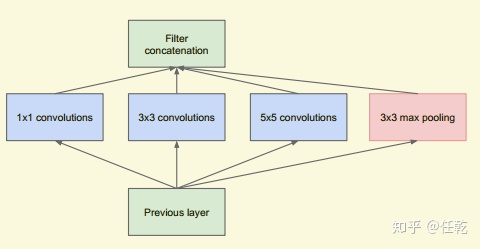
**十、Convolutional Neural Network（CNN）：**

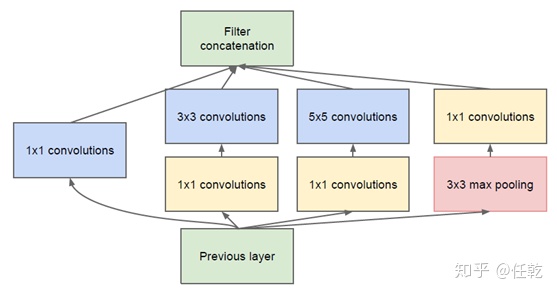
1、AlexNet

2、ZFNet：反卷积，探测卷积网络 反池化、反激活、反卷积

3、VGGNet：使用多个小卷积核构成的卷积层代替较大的卷积层，两个3x3卷积核的堆叠相对于5x5卷积核的视野，三个3x3卷积核的堆叠相当于7x7卷积核的视野。一方面减少参数，另一方面相当于进行了更多的非线性映射，可以增加拟合能力。

4、Googlenet：inception, 提供多个不同大小的卷积核让机器自己选择**（类似于kernel learning？）**





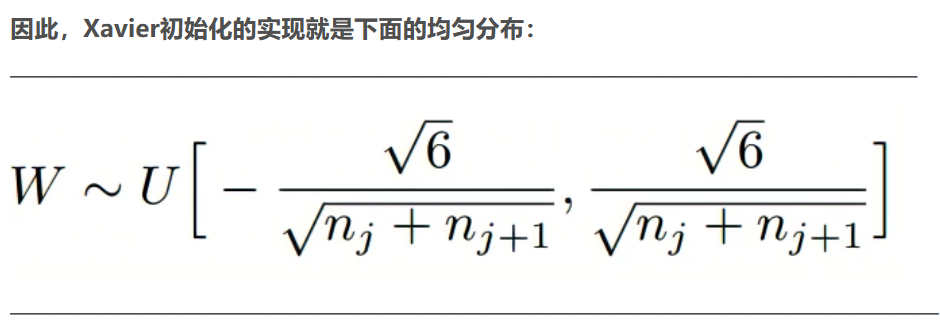
**（为什么下图中增加了1\*1的卷积核会大大减少了参数量？）**

5、Resnet：发现degradation problem，引入残差学习

6、SENet：SE模块

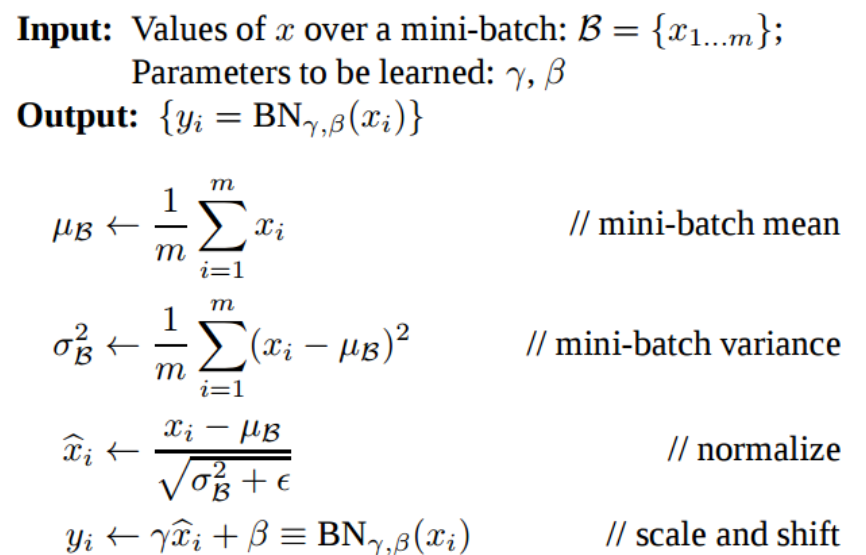
**十一、Weight initialization：**

1、“Xavier”：

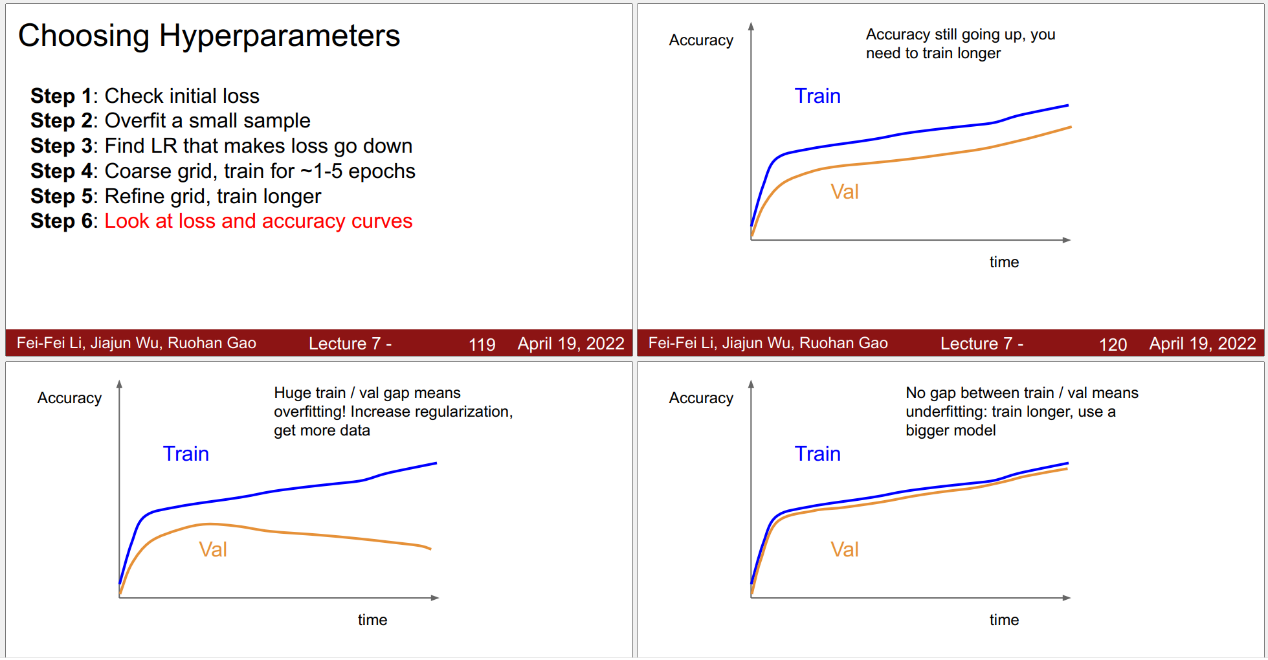


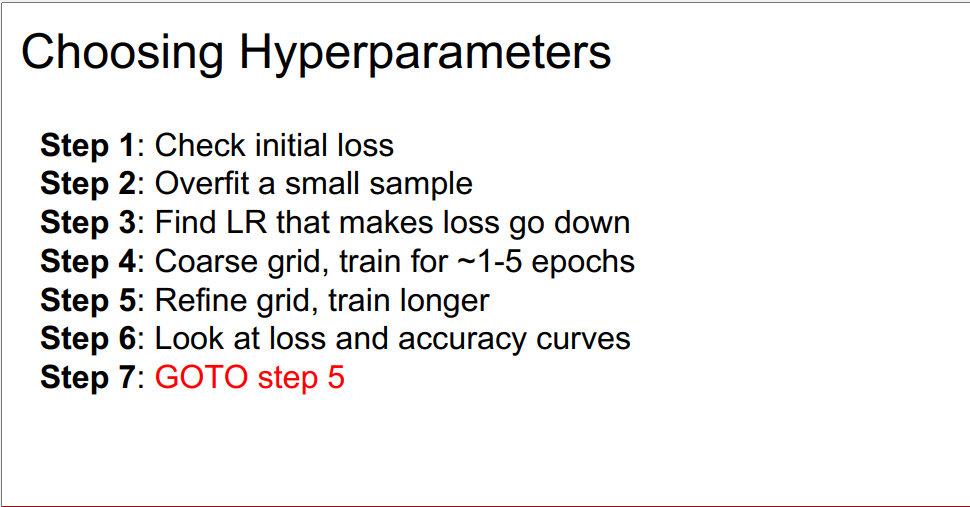
2、He initialization：在ReLU网络中，假定每一层有一半的神经元被激活，另一半为0

3、Batch Normalization Layer：



**十二、Choosing hyperparameters：**





**十三、Visualizing Convnet layers：**

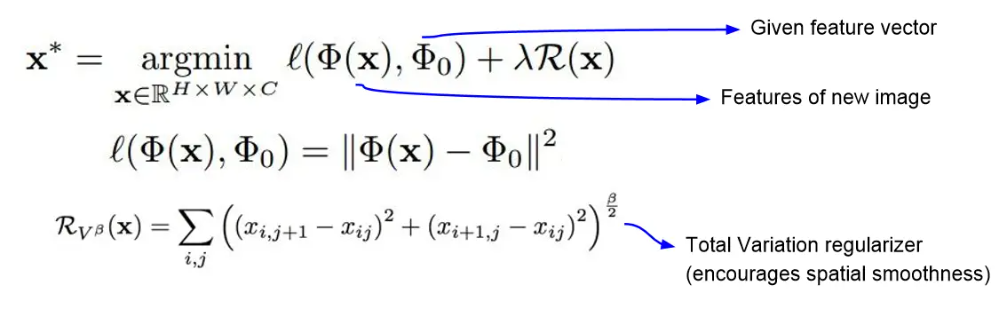
1、First Layer: Visualize Filters

2、Last Layer: Nearest Neighbors、Dimensionality Reduction

3、Maximally Activating Patches（最大激活块）

**十四、Style transfer：**

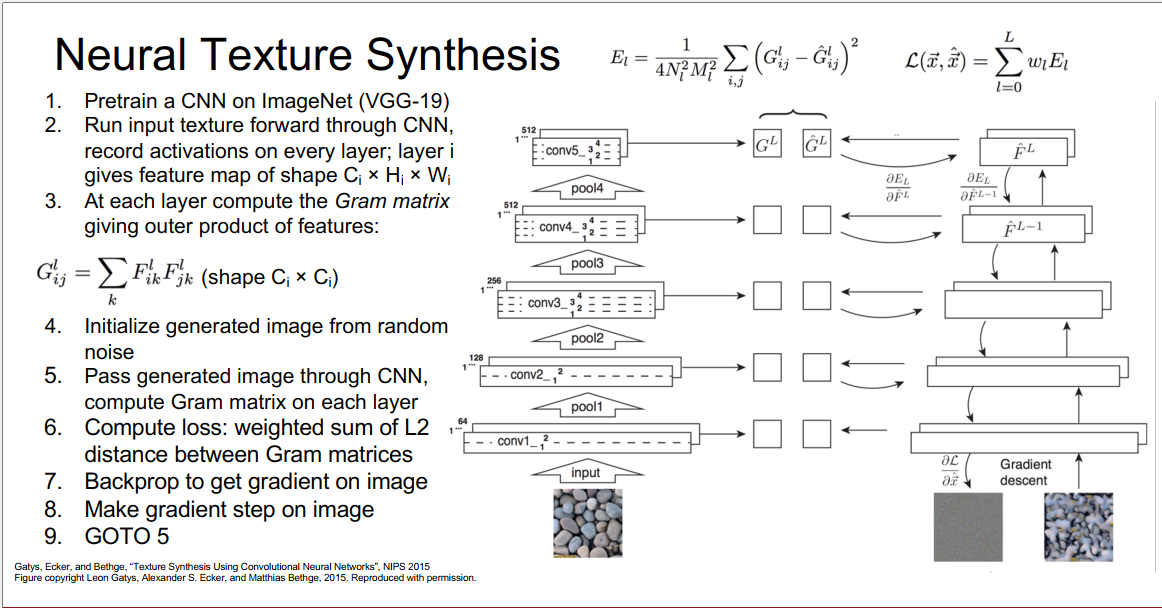
1、Feature Inversion（特征反演）：



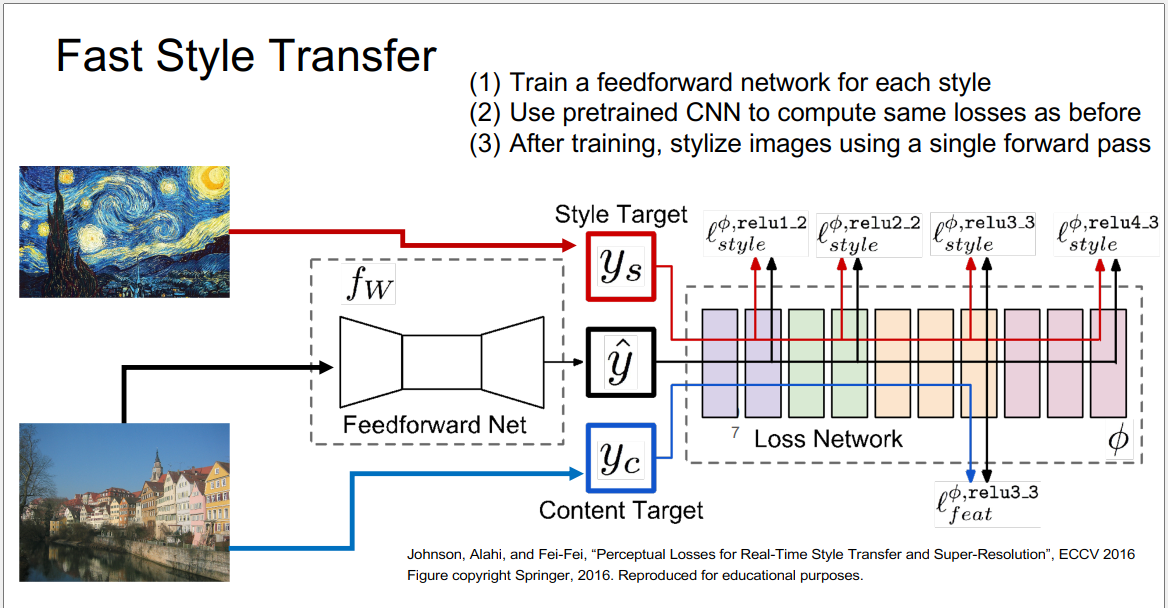
2、Deep Dream

3、Texture Synthesis：Gram Matrix（偏心协方差矩阵）

4、Neural Texture Synthesis：类似梯度上升



5、Fast style transfer：



**十五、Semantic Segmentation：**

1、Fully Convolutional Networks（FCN）: 解决边缘信息不足

2、Learnable Upsampling：

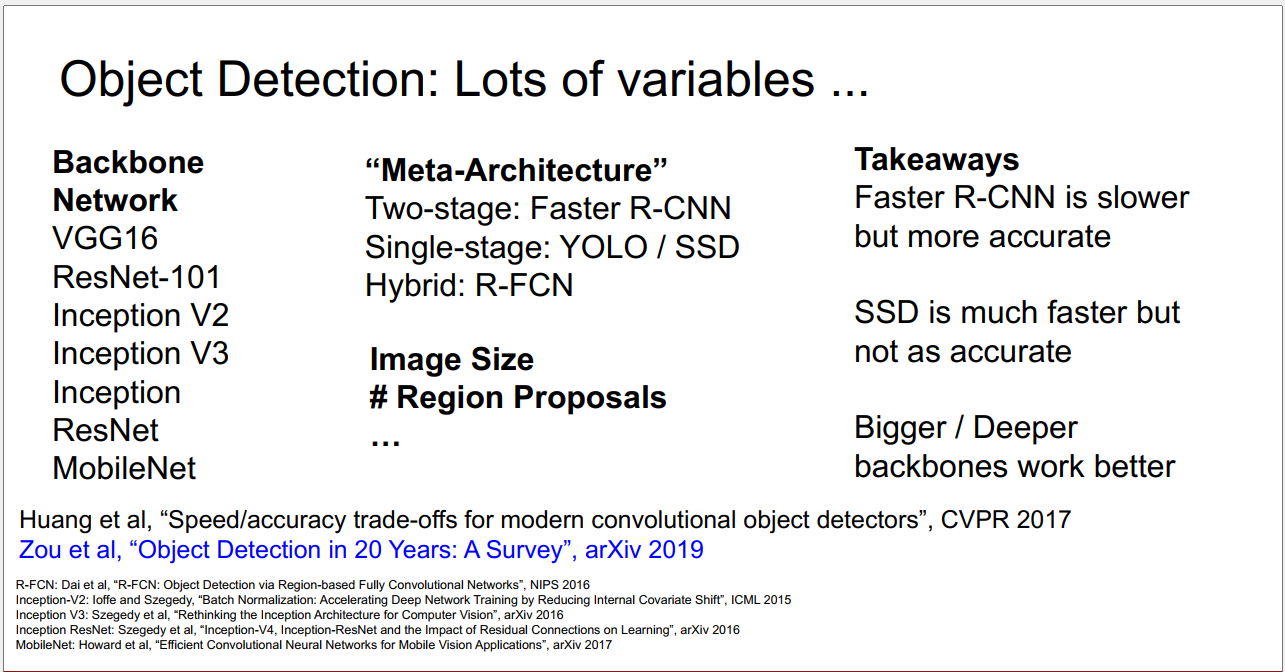
a) 线性插值、双线性插值

b) Transposed Convolution / Deconvolution

**十六、Object detection：**

Region Proposals、Selective Search

1. R-CNN（SVM分类）
2. Fast R-CNN（加入RoI Pooling——Region of interest pooling， Region features slightly misaligned） （ROI Pooling **->** ROI Align）
3. Region Proposal Networks（RPN）



**十七、Instance segmentation：**

1、Mask R-CNN

**十八、Recurrent Neural Networks：**

一个序列当前的输出与前面的输出也有关

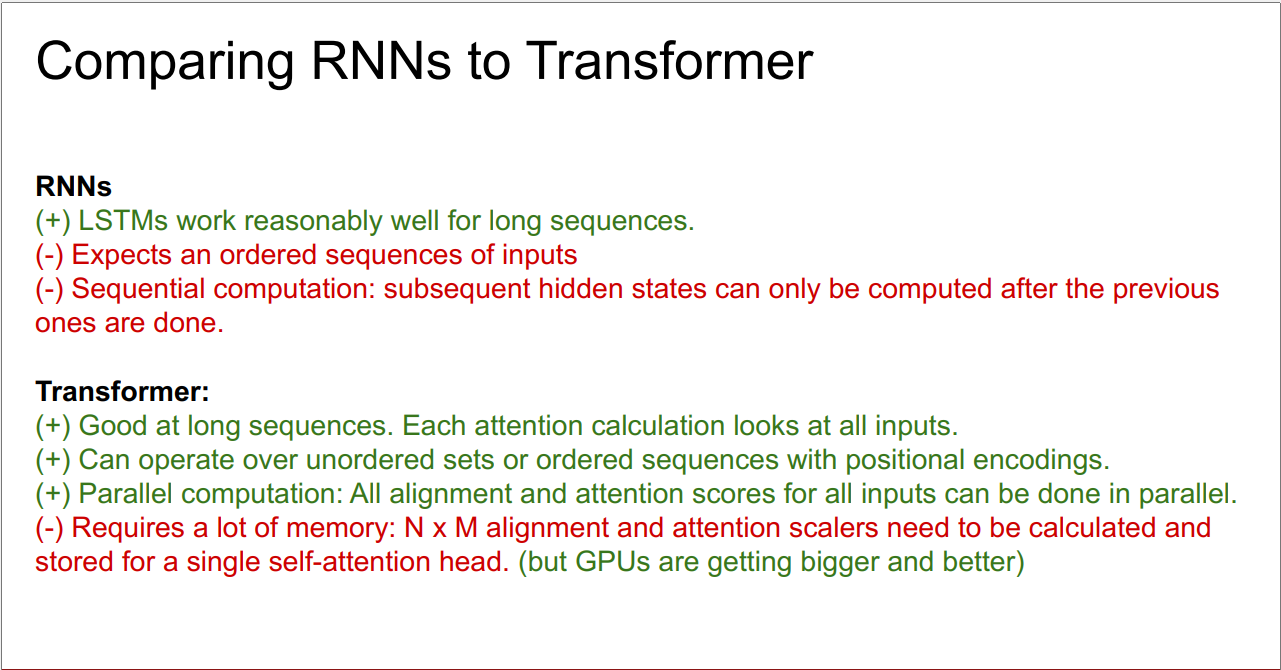
In CV：

1、Image Captioning：CNN+LSTM **->** Faster-RCNN+双层LSTM+注意力：General attention layer、Self attention layer、Positional encoding

2、Long Short Term Memory（LSTM）：

Visual Question Answering（VQA）

Transformers: self-attention + layer normalization （quickly replace RNNs, LSTMs）



**十九、Video classification:**

1、Early Fusion

2、Late Fusion

3、3D CNN

4、Measuring Motion:

a) Optical Flow

b) Two-Stream Networks

5、Spatio-Temporal Self-Attention（STAN）

6、Inflating 2D Networks to 3D（I3D）

7、Spatio-Temporal Detection

**二十、Unsupervised Learning（no labels）：**

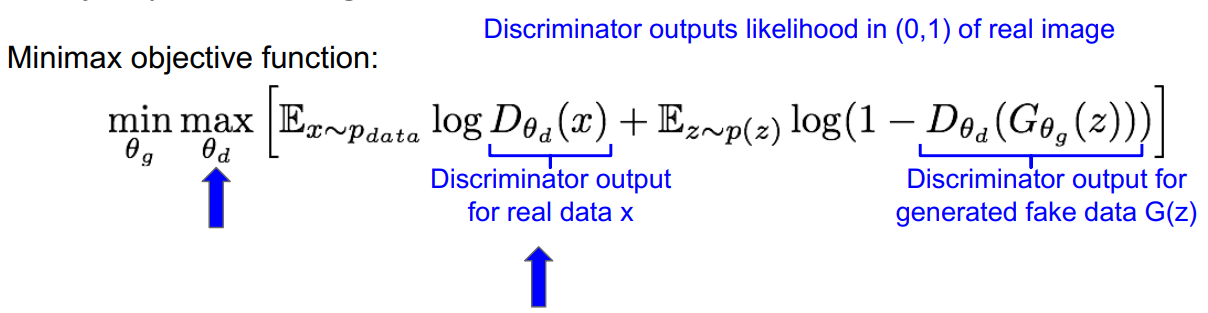
1、Generative Models：X的数据分布/密度估计是无监督学习的核心问题，生成式模型可以解决密度估计问题

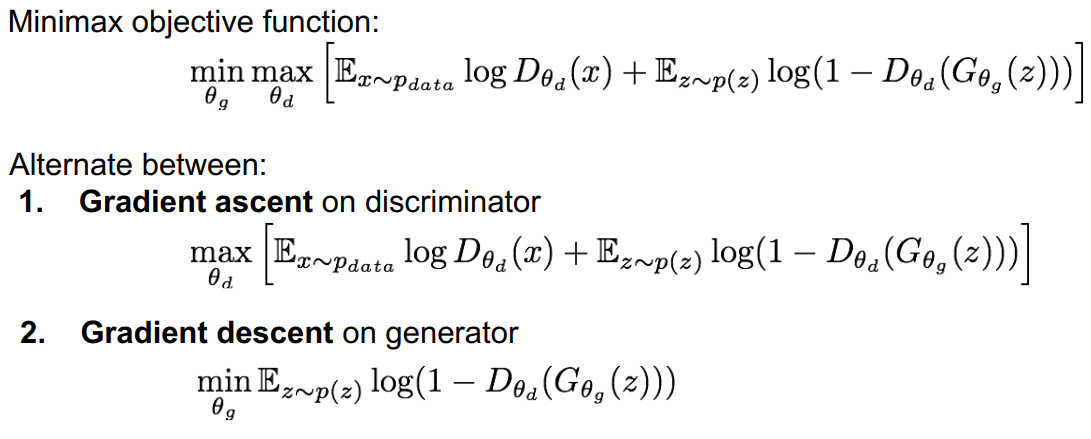
2、Fully visible belief network：（FVBN）

3、PixelRNN、PixelCNN （可以显式计算x的似然概率p(x)，但像素是顺序生成，模型比较慢）

4、Variational Autoencoders（VAE）：

5、Generative Adversarial Networks（GAN）：生成的样本最佳，但训练不稳定，难以训练；不能直接进行推断查询p(x)，p(z | x)

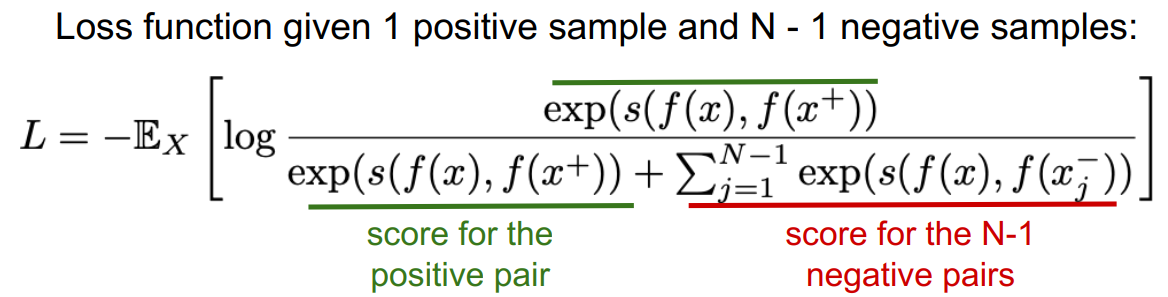




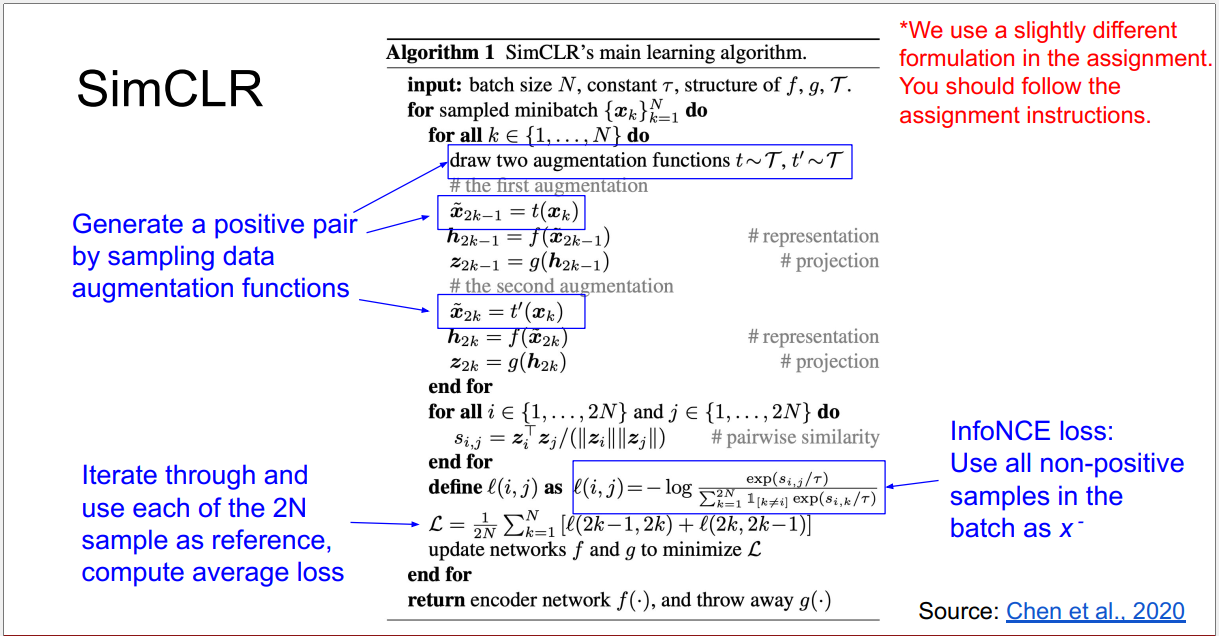
6、Self-supervised Learning：

a) Pretext task: predict rotations、predict relative patch locations、solving “jigsaw puzzles”、 predict missing pixels (inpainting)、image coloring （Problems: coming up with individual pretext tasks is tedious ; the learned representations may not be general.）

7、Contrastive Representation Learning：



1. SimCLR：non-linear projection、large training batch size is better



b) Momentum Contrastive Learning（Moco）: Decouples negative sample size from minibatch size

c) Moco v2: combines nonlinear projection head, strong data augmentation, with momentum

d) Contrastive Predictive Coding（CPC）：sequence-level contrastive learning、InfoNCE loss with a time-dependent score function

**想法：**

1、现已有从结果归纳解释原因的方法：

找到关键像素：

Saliency Maps

Occlusion Experiments

Intermediate Features via (guided) backprop

Gradient Ascent

但仍然属于归纳法，而非演绎法。模型的严格数学证明/可解释性？

2、模型最终绕不开损失函数、激活函数，在基础上的改进会进一步推动现有模型的进展？

3、在数学基础我上还需要加强，尤其是凸优化中的内容。